

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 16 日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730075

研究課題名(和文) 広域仮想クラスタのためのネットワークスループットアウェアな仮想ネットワークの研究

研究課題名(英文) A study on throughput-aware virtual network for widely distributed virtual clusters

研究代表者

市川 昊平 (Ichikawa, Kohei)

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号：90511676

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究では、ソフトウェアによってネットワークを制御する技術であるSDNを活用し、拠点間ネットワークの帯域や遅延を考慮に入れ自動的に最適な仮想ネットワークを動的に構成する、ネットワークスループットアウェアな仮想ネットワーク構成技術の開発を目的とし、それを実現するための要素技術の開発および実証的実験を実施した。提案手法を適用することによって、従来手法に比べネットワークのパフォーマンスが大幅に改善されることを示した。

研究成果の概要(英文)：This study has developed fundamental technologies and performed experiments using SDN to build a throughput-aware virtual network system which allows us to automatically deploy optimal virtual networks in terms of network bandwidth and latency. This study figured out that the proposed method improved the performance of network compared to the conventional method.

研究分野：分散システム

キーワード：SDN ネットワークスループットアウェア 仮想ネットワーク 仮想クラスタ

1. 研究開始当初の背景

仮想計算機(VM)技術の発達により、計算資源の配備はソフトウェアによって動的に制御可能となり、計算資源の共有・利用効率を高め、米国 Amazon 社における Amazon EC2 に代表されるようなスケラブルで柔軟な仮想計算機を提供するクラウドサービスを実現可能としている。計算科学分野においても、大規模なデータ解析やシミュレーションを実行する基盤環境を柔軟に構築できる技術として、動的に配備した仮想計算機を複数台集約して仮想的なクラスタを構築する技術が注目を集めている。さらに、今日では、地理的に分散する複数の研究機関や大学、データセンタの保有する計算資源上に、動的に仮想計算機を配備し、仮想ネットワーク技術等により集約することによって、複数拠点間にまたがるマルチサイト仮想クラスタを構築する技術に関する研究開発が推進されつつある(図1)。マルチサイト仮想クラスタでは、既存の計算科学アプリケーションを修正することなく、複数のサイトにまたがる計算資源を有効に活用可能であるため、大きく期待が高まっている。

このように計算資源割り当てを動的にソフトウェアで制御する技術に関しては成熟しつつある一方で、ネットワークの制御技術は未確立な現状がある。マルチサイト仮想クラスタを構成する上で、拠点間で仮想ネットワークを構築する目的は、拠点間のネットワークポリシーの相違や物理トポロジを隠蔽し、透過的な単一ネットワークを仮想計算機に提供することで計算機間の相互接続性を確保することにある。しかし、その一方で、仮想ネットワークは物理ネットワークのト

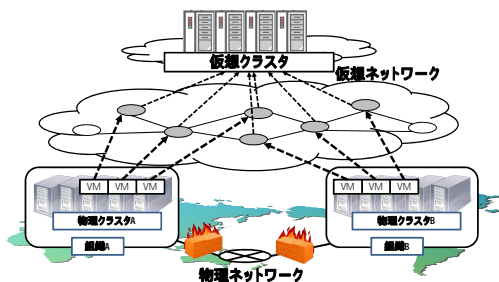


図2 マルチサイト仮想クラスタの概念図

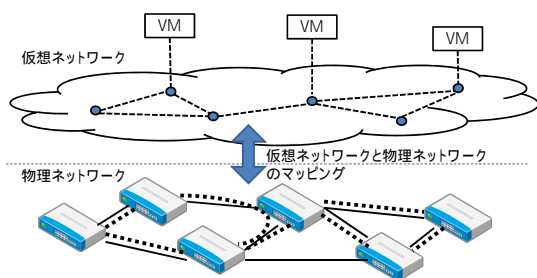


図1 仮想ネットワークと物理ネットワークのマッピングの相違

ポロジを隠蔽してしまうため、計算機側から各仮想ネットワークリンクの帯域や遅延などを把握することが困難であり、性能を引き出すための最適なルーティングができない問題がある。これは図2に示すように、計算機に提供される論理的な仮想ネットワークトポロジからは、それが実際に物理ネットワークにどのようにマッピングされているかわからないことに起因する。したがって、仮想ネットワーク上では最短経路でルーティングしているつもりであっても、実際の物理的なルーティングは必ずしも最適ではない場合が生じる。マルチサイト仮想クラスタにおいて、仮想ネットワークはインターコネクトネットワークとして使用されるため、その性能上の問題は仮想クラスタ全体の性能や応答性に多大な影響を与える。したがって、仮想クラスタを実用的に活用可能とするためには、この問題の解決は不可欠であると考えられる。

このような背景から、本研究実施者は物理ネットワークのトポロジ、ネットワーク帯域、遅延を考慮に入れ、自動的に最適なルーティングを実施するネットワークスループットウェアな仮想ネットワークの実現が急務であると考え、本研究の着想に至った。本研究実施者はこれまでの研究で、マルチサイト仮想クラスタの実現に向けて、P2P 技術ベースのオーバーレイネットワーク上で仮想ネットワークを構成する手法の開発や、近年では SDN (Software Defined Networking) 技術を活用した仮想ネットワーク構成手法を研究開発し、国際的な広域ネットワーク上で実証実験を推進してきた。実証実験において SDN を活用した理由は、異なるユーザの通信を隔離し独立性を確保することが目的であったが、SDN が可能とするソフトウェアによるネットワークの柔軟な制御によって、本研究で求められる最適ルーティングを実現できる可能性を感じたことも、本研究の着想に至った理由の一つである。

2. 研究の目的

本研究では、ソフトウェアによってネットワークを制御する技術である SDN を活用し、拠点間ネットワークの帯域や遅延を考慮に入れ自動的に最適な仮想ネットワークを動的に構成する、ネットワークスループットウェアな仮想ネットワーク構成技術の開発を目的とし、それを実現するための要素技術の開発および実証の実験を実施した。

3. 研究の方法

本研究は、本研究実施者がこれまでに構築してきた SDN による仮想ネットワークを発展する形で実施した。具体的には、SDN のための代表的な標準プロトコルである OpenFlow を活用し、最適ルーティングのためのネットワーク制御を広域ネットワーク環境で実施することとした。ただし、OpenFlow 自体は、

ネットワークを制御するコントローラとネットワークスイッチ間の制御プロトコルを標準化しているだけにすぎない。そのため、本研究で求められるような、自動最適化機能は実装するためには、1) ネットワークトポロジ・帯域・遅延などネットワークの状況を適切に把握し、2) その情報に基づいた最適ルーティングを実施する仮想ネットワークを自動かつ動的に構成する仕組みをOpenFlowのフレームワーク上に実現する必要がある。

そこで、本研究では、本目的を達成するため、マイルストーンとなる以下の3つの課題を設定して、実施してきた。

[課題 1]: ネットワーク状況を収集するネットワークセンサエージェントの設計と実装

提案手法を実現する上で、ネットワークトポロジ・帯域・遅延状況を適切に把握するための情報収集の仕組みは重要である。本研究では、OpenFlowネットワーク上のトポロジおよび帯域・遅延などのパフォーマンス情報を収集するネットワークエージェントを研究開発した。

[課題 2]: 収集情報に基づいた最適ルーティング制御を実施するコントローラの設計と実装

OpenFlowの大きな特徴は、通信フローが通るネットワーク経路を素早く動的に切り替え可能な点にある。本研究では、センサエージェントから得た情報を基に、動的に最適化を実施し、ネットワーク経路を制御するコントローラを実装した。本研究では得られるパフォーマンスを最大化するため、ネットワーク上の複数のパスを同時に割り当て・使用可能とするコントローラに関する研究開発も実施した。

[課題 3]: 提案仮想ネットワーク上におけるマルチサイト仮想クラスタの実証実験による評価

本研究で提案した仮想ネットワークを、実際のマルチサイト仮想クラスタ構築に適用し、広域環境上にて計算科学アプリケーション動作の実証実験を行い、その実用性・有用性について評価を行った。

#### 4. 研究成果

本研究では、前節で挙げた研究目的を達成するために解決が必要な各課題に応じて、要素技術の開発や研究を実施し、それぞれ成果を達成してきた。以下、各項目にしたがって、説明する。

(1) ネットワーク状況を収集するネットワークセンサエージェントの設計と実装

本研究では、OpenFlowネットワークのパフォーマンス状況を半リアルタイムで把握するためのネットワークセンサエージェント

を開発した。本研究ではネットワークの最適化に特化して、ネットワークをモニタリングすることから、リアルタイムに近い情報把握を必要であった。そこで、本研究ではセンサエージェント自身が能動的にネットワーク上に通信を発生させ遅延や利用可能な帯域を測定するアクティブ測定方式も用いることにした。ただ、アクティブ方式にすることで正確に測定時点でのパフォーマンス情報を得られる一方、ネットワークの測定のためにセンサエージェントから発生させる通信同士が衝突することで適切な測定結果が得られない場合があるという課題が生じた。本研究ではこのような問題を回避するため、センサエージェントが動作するホストのCPUやネットワークの負荷をモニタリングし、ネットワーク測定に影響が生じないタイミングおよび測定ホストの組み合わせを探しだし、モニタリングのための通信発生をスケジュールするモニタリングフレームワークを構築した。

(2) 収集情報に基づいた最適ルーティング制御を実施するコントローラの設計と実装

マルチサイトクラスタを利用するアプリケーションにおいては、それぞれネットワークに対する要求が異なる場合がある。あるアプリケーションは大量のデータを送信するため、帯域が大きな回線を好む場合があるし、あるアプリケーションはデータサイズは小さいが非常に頻繁に情報を交換するため、遅延ができるだけ小さい回線を好む場合がある。

そこで、本研究では各アプリケーションの要求に基づいて、そのそれぞれの通信パフォーマンスを最大化可能なネットワークパスを前述のネットワークパフォーマンス情報を基に計算し、アプリケーションごとに割り当てるOpenFlowコントローラを構築した。具体的には前述したネットワークパフォーマンス情報を用いてネットワークリンクごとのコストを帯域および遅延に応じて計算し、アプリケーションの要求に応じて帯域を最大化させることが可能となるパスや、遅延を最小化させるパスを割り当てる制御を行った。その結果、従来のネットワークルーティングでは非効率であったサイト間通信が飛躍的に向上できることを確認した。

また、特に大量データを送信するアプリケーションに関しては、複数のTCPコネクションを利用してデータ送信を行うものが存在するが、本研究ではそのようなアプリケーションの複数TCPコネクションに対し、異なるネットワーク経路を同時に割り当てることで、使用可能な総帯域を大幅に引き上げることも成功した。

(3) 提案システムの実証実験

本研究では、実証実験として、グリッド及びクラウドコンピューティングの国際的な

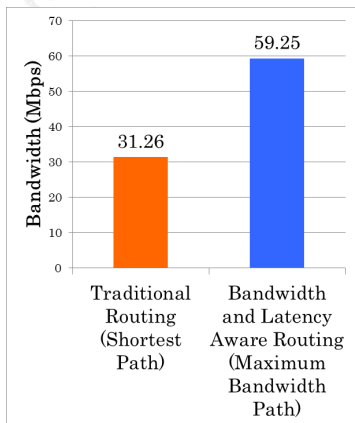
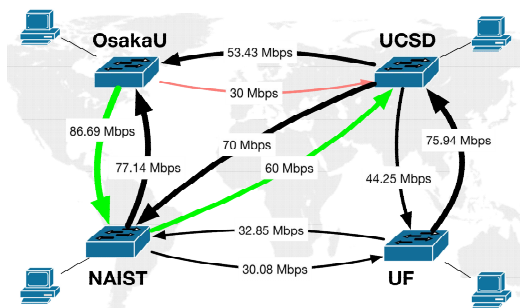


図 3 マルチサイト間における帯域最適化の評価実験結果

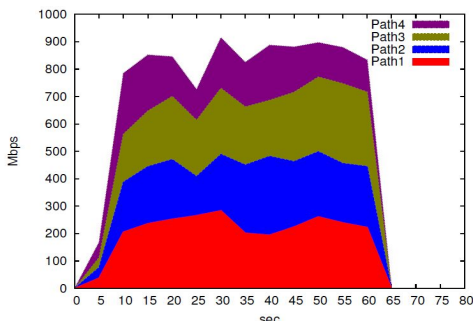


図 4 複数ネットワークパスを用いた帯域最適化の結果

連携コミュニティである PRAGMA の参画機関が提供する計算資源上に広域の OpenFlow ネットワークを構築し、その上に提案システムを配備することで、マルチサイト環境におけるネットワーク性能の最適化に関する評価を行った。

図 3 は、奈良先端科学技術大学院大学、大阪大学、カリフォルニア大学サンディエゴ校、フロリダ大学を結んで構築したネットワーク上で提案システムによりネットワーク帯域を最適化した結果を示す。評価のグラフが示す通り、従来手法に比べて提案手法では倍近くのパフォーマンスを達成することができている。

図 4 は同じく広域 OpenFlow ネットワーク環境において、異なる複数のネットワークパスを同時に用いて最大使用可能帯域を向上させた結果である。本実験を実施した際のクライアント機器に搭載されていたネットワークカードは 1Gbps の製品であり、図 4 の結

果はその上限に近い性能を広域環境下で達成していることを示している。これらの結果は PRAGMA が実施する国際ワークショップにおいて、ライブデモンストレーション等により実証した。

## 5. 主な発表論文等

〔雑誌論文〕(計 1 件)

渡場康弘, 木戸善之, 伊達進, 阿部洋丈, 市川昊平, 山中広明, 河合栄治, 竹村治雄, “計算資源とネットワーク資源を考慮した割当ポリシーを配備可能とするジョブ管理フレームワーク,” 電子情報通信学会論文誌, vol. J97-D, No. 6, pp. 1082-1093, June 2014.

〔学会発表〕(計 12 件)

Che Huang, Chawanat Nakasan, Kohei Ichikawa, Hajimu Iida, “A multipath controller for accelerating GridFTP transfer over SDN,” In 11th IEEE International Conference on eScience, pages 439-447, September 2015, Munich, (Germany).

Yasuhiro Watashiba, Susumu Date, Hirotake Abe, Yoshiyuki Kido, Kohei Ichikawa, Hiroaki Yamanaka, Eiji Kawai, Shinji Shimojo and Haruo Takemura, “Performance Characteristics of an SDN-enhanced Job Management System for Cluster Systems with Fat-tree Interconnect,” Emerging Issues in Cloud Workshop, 6th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2014), December 2014, Singapore (Singapore).

Pongsakorn U-chupala, Kohei Ichikawa, Hajimu Iida, Nawawit Kessaraphong, Putchong Uthayopas, Susumu Date, Hirotake Abe, Hiroaki Yamanaka, Eiji Kawai, “Application-Oriented Bandwidth and Latency Aware Routing with OpenFlow Network,” Emerging Issues in Cloud Workshop, 6th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2014), pp. 775-780, December 2014, Singapore (Singapore).

Yoshiyuki Kido, Kohei Ichikawa, Yasuhiro Watashiba, Hirotake Abe, Hiroaki Yamanaka, Eiji Kawai, Haruo Takemura, Shinji Shimojo, “SAGE-based Tiled Display Wall Enhanced with Dynamic Routing Functionality Triggered by User Interaction,” Innovating the Network

for Data Intensive Science workshop (INDIS2014), The International Conference for High Performance Computing, networking, Storage and Analysis (SC14), November 2014, New Orleans (USA).

Pongsakorn U-chupala, Kohei Ichikawa, Putchong Uthayopas, Susumu Date, Hirotake Abe, "Designing of SDN-Assisted Bandwidth and Latency Aware Route Allocation," 研究報告ハイパフォーマンスコンピューティング (HPC), Vol. 2014-HPC-145, No. 2, pp. 1-7, 2014年7月, 朱鷺メッセ新潟コンベンションセンター (新潟県新潟市).

Chawanat Nakasan, Kohei Ichikawa, Putchong Uthayopas, "Performance Evaluation of MPTCP over OpenFlow Network," 研究報告ハイパフォーマンスコンピューティング (HPC), Vol. 2014-HPC-145, No. 30, pp. 1-6, 2014年7月, 朱鷺メッセ新潟コンベンションセンター (新潟県新潟市).

黄 掣, 市川 昊平, "SDNによるマルチパスを用いた GridFTP によるデータ転送高速化手法の提案," 研究報告ハイパフォーマンスコンピューティング (HPC), Vol. 2014-HPC-145, No. 1, pp. 1-8, 2014年7月, 朱鷺メッセ新潟コンベンションセンター (新潟県新潟市).

Nawawit Kessaraphong, Putchong Uthayopas and Kohei Ichikawa, "Building a Network Performance Benchmarking System Using Monitoring as a Service Infrastructure," The 18th International Computer Science and Engineering Conference (ICSEC2014), July 2014, Khon Kaen (Thailand).

Yasuhiro Watashiba, Yoshiyuki Kido, Susumu Date, Hirotake Abe, Kohei Ichikawa, Hiroaki Yamanaka, Eiji Kawai, Haruo Takemura, "Prototyping and Evaluation of a Network-aware Job Management System on a Cluster System Leveraging OpenFlow," The 19th IEEE International Conference On Networks (ICON 2013), December 2013, Singapore (Singapore).

Kohei Ichikawa, Susumu Date, Hirotake Abe, Hiroaki Yamanaka, Eiji Kawai and Shinji Shimojo, "A Network Performance-Aware Routing for Multisite Virtual Clusters," The 19th IEEE International Conference on Networks (ICON2013), December 2013, Singapore (Singapore).

Yasuhiro Watashiba, Susumu Date, Hirotake Abe, Kohei Ichikawa, Hiroaki Yamanaka, Eiji Kawai, Haruo

Takemura, "An Architectural Design of a Job Management System Leveraging Software Defined Network," The 4th IEEE International Workshop on High-Speed Network and Computing Environment (HSNCE 2013), pp.724-729, July 2013, Kyoto Terrsa (Kyoto, Japan).

市川 昊平, 伊達進, 阿部洋丈, 渡場康弘, 下條真司, "SDN 技術を用いたインターネットクラウド環境における分散計算効率化に関する研究," 第2回地域間インターネットワークショップ, 2013年5月, 金沢大学 (石川県金沢市).

#### [ 図書 ] (計 1 件)

Yoshio Tanaka, Naotaka Yamamoto, Ryousei Takano, Akihiko Ota, Philip Papadopoulos, Nadya Williams, Cindy Zheng, Weicheng Huang, Yi-Lun Pan, Chang-Hsing Wu, Hsi-En Yu, J.H. Steven Shiao, Kohei Ichikawa, Taiki Tada, Susumu Date, and Shinji Shimojo. Building secure and transparent inter-cloud infrastructure for scientific applications. In C. Catlett, W. Gentzsch, L. Grandinetti, G.R. Joubert, and J.L. Vazquez-Poletti, editors, Cloud Computing and Big Data, volume 23 of Advances in Parallel Computing, pages 35-52. IOS Press, October 2013.

#### 6. 研究組織

##### (1) 研究代表者

市川 昊平 (ICHIKAWA, Kohei)

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号 : 90511676

##### (2) 研究分担者

なし

##### (3) 連携研究者

なし